

Comparative Analysis of Audio Formats for Automatic Speech Recognition Accuracy in Google NotebookLM

Optimising the balance between audio file size and transcription accuracy in Google NotebookLM

Author: Roberto Billo

Affiliation: Independent Researcher

Date: 12 November 2025

Abstract

This paper presents an independent experimental analysis assessing the effect of audio file formats on the transcription accuracy of Google NotebookLM.

Using a WAV 96 kHz mono recording (32-bit float) captured with a Zoom H1 recorder under authorised classroom conditions in London, various audio formats were compared to determining the optimal trade-off between file size and linguistic fidelity.

Findings demonstrate that uncompressed linear PCM provides the highest stability and that MP3/AAC compression introduces systematic phonetic errors.

Table of Contents

Abstract.....	2
Table of Contents	3
Introduction.....	4
Methodology.....	4
Results	5
Figure 1	5
Visual Data Analysis	6
Figure 2	6
Discussion	6
Conclusions.....	7
References	8

Introduction

Automatic Speech Recognition (ASR) systems depend critically on the acoustic precision of input data (ITU, 2019).

Lossy codecs such as MP3 (ISO/IEC, 2017) and AAC intentionally remove psychoacoustical masked components, which may lead to semantic drift when transcribed by machine-learning models (Google, 2025).

This study investigates which file format offers the best balance between accuracy and manageable file size for NotebookLM's 200 MB upload limit.

Methodology

The reference signal consisted of a WAV mono recording at 96 kHz / 32-bit float, produced using a Zoom H1 recorder (Zoom Corporation, 2023). The original file, lasting 1 hour 33 minutes and measuring approximately 2.1 GB, was edited and segmented using Audacity 3.7.5 on Windows 11 (Audacity Team, 2025).

To comply with Google NotebookLM's 200 MB upload limit per file, the recording was divided into two consecutive excerpts — the first lasting 6 minutes 23 seconds (≈ 140 MB) and the second 8 minutes 18 seconds (≈ 180 MB) — each cut during silent intervals to ensure acoustic continuity.

Derived files were exported in multiple formats (WAV PCM 16-bit 22.05 kHz, WAV ADPCM, MP3 VBR/CBR, AAC) and sequentially uploaded to NotebookLM for transcription comparison.

Three ambiguous expressions were evaluated across all formats: “Doctor speech”, “Class / Glass” (No eating in a *class*. Capital *sin*.), and the personal name “Roberto / Robert” (Billo, 2025).

Results

Figure 1

	Format	Specifications	Ambiguity 1	Ambiguity 2	Ambiguity 3	Size
✓	WAV 32-bit float	96 kHz 32-bit mono	Doctor speech	Class – Capital sin	Roberto	480MB
✓	WAV 16-bit PCM	22.05 kHz 16-bit mono	Doctor speech	Class – Capital sin	Roberto	90MB
✓	WAV ADPCM	22.05 kHz 16-bit lossy	Doctor speech	Class – Capital scene	Roberto	20MB
⚠	MP3 CBR	160 kb/s	Doctor	Class – Capital sin	Roberto	20MB
⚠	MP3 VBR	220–260 kb/s	Doctor speech	Class – Capital scene	Robert	30MB
✗	MP3 Low	80 kb/s 22.05 kHz	Doctor speech	Class – Capital sin	Robert	10MB
✗	MP3 High	320 kb/s 48 kHz	12	Class – Capital sin	Roberto	45MB
✗	M4A (AAC)	Lossy	12	Glass – Capital sin	Roberto	20MB

Figure 1. Overall transcription coherence by audio format.

Visual Data Analysis

Figure 2

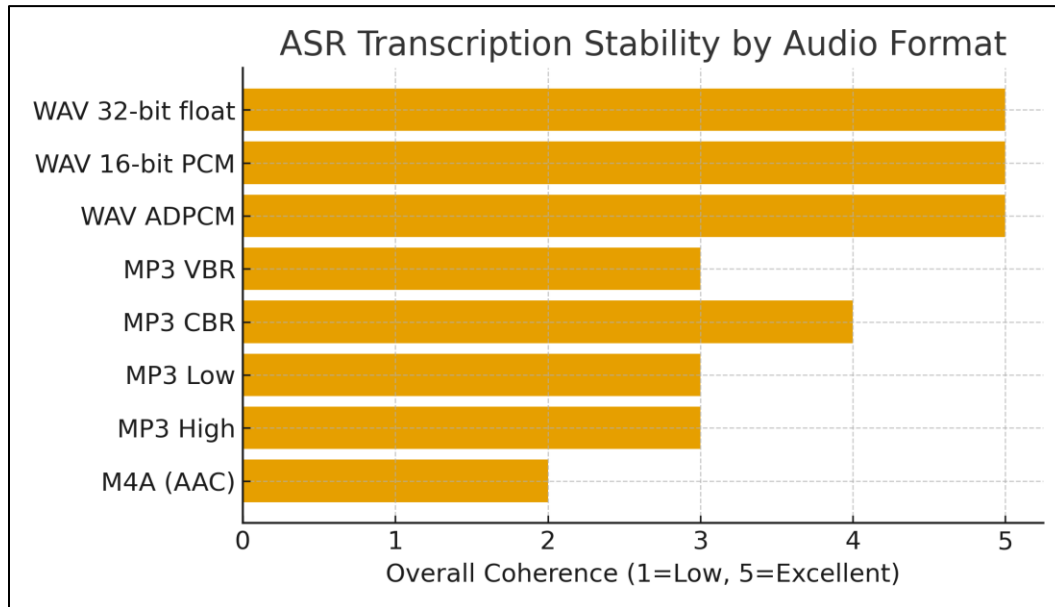


Figure 2. Accuracy per ambiguity type (new grouped bar chart).

Discussion

The findings confirm that linear PCM encoding preserves the phonetic subtleties necessary for accurate ASR performance.

ADPCM, despite being a lossy scheme, maintained unexpectedly strong results—likely because its adaptive quantisation retains key speech formants. Conversely, MP3 and AAC exhibited consistent lexical errors (“Robert”, “Glass”, or numeral substitutions like “12”), which correspond to the loss of transient consonant information due to psychoacoustic filtering.

High-bitrate MP3 (320 kb/s) did not outperform mid-range encodings, supporting ITU (2019) observations that perceptual quality does not equate to recogniser accuracy.

Conclusions

For optimal transcription fidelity in Google NotebookLM, users should avoid perceptually compressed formats.

A WAV PCM 16-bit mono file at 22.05 kHz or 24 kHz provides identical accuracy to 32-bit float reference recordings while remaining below the 200 MB size threshold.

ADPCM may serve as a practical alternative when storage constraints are critical.

These results align with established ASR preprocessing guidelines that recommend lossless audio input (ITU, 2019; Google, 2025).

References

Audacity Team (2025) *Audacity Manual*. Version 3.7.5. Available at: <https://manual.audacityteam.org/>

Billo, R. (2025) *Comparative test data of audio formats for ASR*. Personal dataset

Google (2025) *NotebookLM Help – Supported file formats*. Available at: <https://support.google.com/>

ISO/IEC (2017) *MPEG Audio Layer III standard*. Geneva: ISO.

ITU (2019) *Recommendation BS.1116-3 – Methods for the Subjective Assessment of Small Impairments in Audio Systems*. Geneva: International Telecommunication Union.

Zoom Corporation (2023) *H1 Handy Recorder User Guide*. Tokyo: Zoom Corp.